

Cars and Gas

Bill Rising
StataCorp LP

2014 Nordic and Baltic Stata Users Group meeting
Aarhus University
05 Sept 2014

Contents

1	Introduction	1
1.1	Introduction	1
2	Basics	2
2.1	Data Management and EDA	2
2.2	Basic Analysis	4
3	Conclusion	5
3.1	Conclusion	5

1 Introduction

1.1 Introduction

Introduction

- This is just a small lesson for illustration
 - ◊ Small is important
 - Most of what we need is already here
 - We'll add some code, add a figure, and move things around
-

Starting Up

- We'll use the auto dataset

```
. sysuse auto
```

(1978 Automobile Data)
 - As we all know, this is a US-centric dataset about old cars
 - We would like to make the dataset metric-friendly
 - We would like to model energy consumption of these ancient cars
-

2 Basics

2.1 Data Management and EDA

Converting the Data to a Reasonable System

- We'd like to make the dataset metric-friendly
- We'll run the code in the lesson, but show it only in the handouts

```
. // Changing mpg to liter/100km
. ** 1 mile = 1.609 km; 1 gallon = 3.7854 liters
. gen lp100km = 100/(mpg*(1.609 / 3.7854))
. label var lp100km "Liters per 100km"
.
. // Trunk size in liters
. rename trunk trunk_cuft
. ** 1 inch = 2.54 cm, 1 foot = 12 inches, 1000 cc = 1 liter
. gen trunk = (trunk_cuft*(12*2.54)^3)/1000
. label var trunk "Trunk capacity (liters)"
.
. // Weight in Kilos
. ** 1kg = 2.205 lb
. rename weight weight_lb
. gen weight = weight_lb / 2.205
. label var weight "Weight in kg"
.
. // Length in cm
. rename length length_in
. gen length_cm = length_in * 2.54
. label var length_cm "Length in cm"
.
. // Turning radius in cm
. rename turn turn_circ_ft
. gen turn_rad = round((turn_circ_ft * 12 * 2.54)/(2*c(pi)))
. label var turn_rad "Turning radius in cm"
.
. // Displacement in liters
. rename displacement displacement_cuin
. gen displacement = displacement_cuin * 2.54^3 / 1000
. label var displacement "Displacement in liters"
```

A Peek at the Dataset

- Here is the structure of our new dataset

```
. describe
```

```
Contains data from /Applications/Stata/ado/base/a/auto.dta
  obs:                74                1978 Automobile Data
 vars:                18                13 Apr 2013 17:45
 size:               4,958              (_dta has notes)
```

```
-----
      storage   display   value
variable name  type    format   label    variable label
-----
make          str18    %-18s           Make and Model
price         int      %8.0gc          Price
mpg           int      %8.0g           Mileage (mpg)
rep78         int      %8.0g           Repair Record 1978
```

headroom	float	%6.1f		Headroom (in.)
trunk_cuft	int	%8.0g		Trunk space (cu. ft.)
weight_lb	int	%8.0gc		Weight (lbs.)
length_in	int	%8.0g		Length (in.)
turn_circ_ft	int	%8.0g		Turn Circle (ft.)
displacement_n	int	%8.0g		Displacement (cu. in.)
gear_ratio	float	%6.2f		Gear Ratio
foreign	byte	%8.0g	origin	Car type
lp100km	float	%9.0g		Liters per 100km
trunk	float	%9.0g		Trunk capacity (liters)
weight	float	%9.0g		Weight in kg
length_cm	float	%9.0g		Length in cm
turn_rad	float	%9.0g		Turning radius in cm
displacement	float	%9.0g		Displacement in liters

Sorted by: foreign

Note: dataset has changed since last saved

- Here are the summary statistics

```
. summarize
```

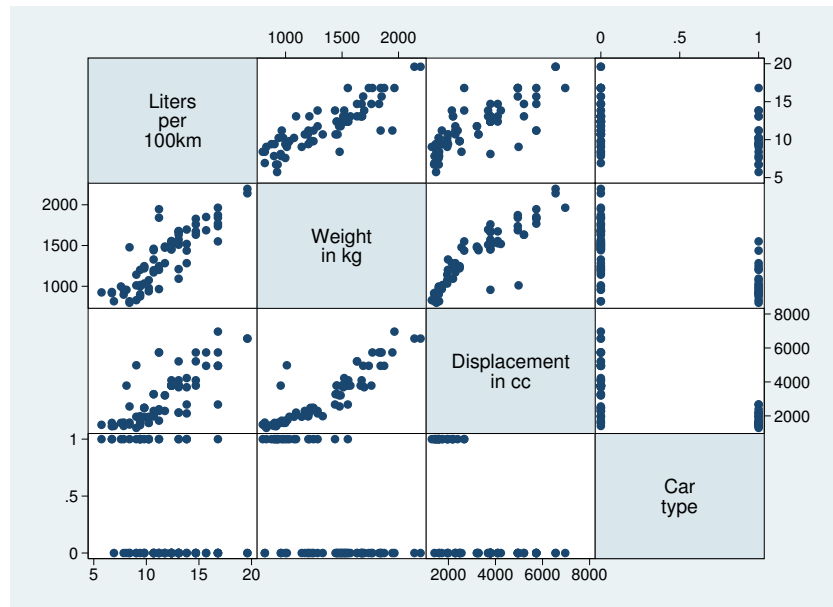
Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74	6165.257	2949.496	3291	15906
mpg	74	21.2973	5.785503	12	41
rep78	69	3.405797	.9899323	1	5
headroom	74	2.993243	.8459948	1.5	5
trunk_cuft	74	13.75676	4.277404	5	23
weight_lb	74	3019.459	777.1936	1760	4840
length_in	74	187.9324	22.26634	142	233
turn_circ_ft	74	39.64865	4.399354	31	51
displaceme-n	74	197.2973	91.83722	79	425
gear_ratio	74	3.014865	.4562871	2.19	3.89
foreign	74	.2972973	.4601885	0	1
lp100km	74	11.80857	3.011042	5.73815	19.60534
trunk	74	389.548	121.1226	141.5842	651.2875
weight	74	1369.369	352.4687	798.1859	2195.011
length_cm	74	477.3484	56.5565	360.68	591.82
turn_rad	74	192.4054	21.33934	150	247
displacement	74	3.233123	1.504942	1.294578	6.964502

- We now should look closer

A Graph Matrix

- We can look at dependencies

```
. graph matrix lp100km weight displacement foreign
```



2.2 Basic Analysis

A Simple Model

- From our graph, it appears we should try

```
. regress lp100km weight displacement foreign
```

Source	SS	df	MS	Number of obs =	74
Model	506.885959	3	168.961986	F(3, 70) =	76.33
Residual	154.95926	70	2.21370372	Prob > F	= 0.0000
Total	661.84522	73	9.06637287	R-squared	= 0.7659
				Adj R-squared	= 0.7558
				Root MSE	= 1.4879

lp100km	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
weight	.007496	.0011156	6.72	0.000	.0052709 .0097211
displacement	.2676906	.2665308	1.00	0.319	-.2638883 .7992695
foreign	1.575823	.4830161	3.26	0.002	.6124776 2.539168
_cons	.2097918	1.019557	0.21	0.838	-1.823651 2.243235

- The coefficient of displacement is 0.26769 with a p -value of 0.31866855

Some Great Explanations

- Heavier cars used more gas
 - After adjusting for engine size and place of manufacture
- Non-US made cars used more gas
 - After adjusting for weight and engine size

- Engine size itself is not important
 - ◊ After adjusting for weight and place of manufacture
-

3 Conclusion

3.1 Conclusion

Conclusion

- These results are interesting
 - We could make the results more interesting with more grant money
-

Index

A

analysis, 2–5

M

metric system, 2